Proceedings of the INMM & ESARDA Joint Virtual Annual Meeting
August 23-26 & August 30-September 1, 2021

A Sound Evaluation of Safeguards Verification Measurements Based on Statistical Analysis and Visualisation of Historical NDA Data

A. Cerasa[1], M. Couland[2], D. Perrotta[1], A. Tomanin[2], C. Versino[3]

[1] European Commission, Joint Research Centre, I.3
[2] European Commission, Directorate-General for Energy, E.1
[3] European Commission, Joint Research Centre, G.II.7

Abstract

The European Commission Directorate-General for Energy performs sets of conformity assessment activities in the form of accountancy and physical verifications in installations using nuclear material. The on-site inspection process makes extensive use of independent measurement systems to determine the flows, quantities or characteristics of nuclear material. The deployment of an increasing number of unattended systems associated with remote data transmission offers an opportunity to improve inspection effectiveness and efficiency by systematic data analysis.

At present, dedicated IT applications are used to evaluate acquired measurement data. The results of their evaluations, being reported by individual inspection, are only representative of a situation over a limited period in time. However, Euratom Safeguards long-term strategy foresees an assessment of risk based on the results of past verification activities and the evaluation of confidence factors capturing limits and uncertainties encountered during the process.

The evaluation of measurement data can directly support those two concepts by including, for instance, a deeper analysis of historical trends or an automated assessment of the measurement systems performances. To this end, the development of dedicated analysis packages associated with a centralised measurement data repository was initiated.

The objective is to provide inspectors with data analysis tools that combine robust statistical techniques and time series analysis aimed to detect anomalous patterns in the measurements. The prompt detection of anomalies such as structural breaks or outliers together with a suitable visualisation of these patterns may indeed support the identification of discrepancies in inspections’ findings. In this paper, we describe an exploratory study of possible analysis tools applied to Non-Destructive Assay data designed to provide a solid basis for the development of a structured and sound statistical framework for the analysis of inspections’ results.

1. Context

The Euratom Treaty laid down the foundation for the peaceful use of nuclear materials and technologies in Member States of the European Union (EU MS). The Treaty established a nuclear material supervision system, known as “Euratom safeguards” to ensure the non-diversion of nuclear material from their intended uses and compliance with safeguards obligations under international agreements.
The European Commission is entrusted with the responsibility of administrating Euratom safeguards and the Directorate-General for Energy, in particular the Directorate E (EC ENER E) implements associated supervision activities by a combination of material accountancy measures and physical verifications.
On-site, Euratom inspectors verify a nuclear facility operator’s declarations related to the flows, quantities and characteristics of nuclear materials with independent findings, these being supported by the analysis of data generated by destructive analysis (DA), containment and surveillance systems and lastly data generated by non-destructive assays (NDA), which are the subject of this paper.
The NDA systems are either operated in attended mode or installed as unattended monitoring equipment. In some cases, NDA systems are coupled with remote data transmission, which offers advantages in terms of implementation of safeguards in nuclear facilities. It allows for a reduction of the level of intrusiveness in the facility operations while increasing the verification coverage, shifting part of the inspection effort from on-site activities to off-site data evaluation.

2. Problem statement

Inspectors typically evaluate NDA results on-site by means of dedicated algorithms running in parallel to the instrument acquisition software. The results of their evaluation are reported by individual inspection and therefore representative of the measurements trend over a limited period in time. In addition, the raw data remain usually on-site and are unavailable for historical analysis or re-evaluation at Euratom Headquarters. Therefore, the conclusion drawn for an inspection tends to not take into consideration trend analysis or cross-cutting different sources of data.
The possibility to perform statistical analysis, in an automated, structured and principled manner and possibly in connection with other existing data handling structures was a clear need identified by Euratom inspectors. It will allow:

- An analysis of the historical trends to complement the data evaluation, for example by definition of validation and decision thresholds based on past results.
- A dynamic assessment of the performance of the measurement system, allowing for maintenance planning and reducing reaction times in case of failures.

On the long term, these two concepts will contribute to an enhancement of the inspection approach. By combining the use of past verification results and the definition of levels of confidence capturing the uncertainties and limits encountered during the process, inspectors will be able to modulate future verification activities more efficiently.
To respond to these needs, the EC ENER E started the implementation of a centralised repository of measurement data coupled with the statistical evaluation of historical results. The European Commission Joint Research Centre (EC JRC) is developing the approach to the statistical analysis based on anonymised datasets of real inspection data.

The following section presents a proposal for the statistical analysis approach. The analysis is based on two anonymised datasets of passive neutron coincidence measurements with different features:

- Use case 1 contains measurement of Plutonium oxide cans with similar characteristics in both content and form. The measurements are taken in unattended mode in combination with a High Purity Germanium Detector to evaluate the isotopic composition of the items.
- Use case 2 contains measurements of impure Plutonium bearings items. With respect to the previous use case, the impurities as well as the packaging of the material add inhomogeneity that affects the measurement uncertainty. The measurements are taken in attended mode during yearly inspection, thus a lower number of measurement points are available for the statistical analysis with respect to use case 1.

Results are presented in terms of potential outliers meaning measurements evaluated as being not coherent with the operator declarations. These are reported along with the outliers identified by inspectors based on their classical evaluation approach to provide a validation of the method proposed. Finally, section 4 describes modalities for embedding the method in the operation workflow of inspections as well as future applications of the statistical analysis.

3. The analytical approach

This section describes the analytical approach based on 3-steps for the analysis of NDA data. The goal is to provide inspectors with data analysis tools aimed at detecting anomalous patterns in NDA measurements, also in relation to the operator's declaration. Considering a simulated sequence $Y_1, \dots, Y_N$ of values for a generic variable Y, Figure 1 shows examples of the anomalous patterns we pursue to identify. A *level shift* is a "jump" in the average level of the historical values of a measured variable. An *outlier*, instead, is a single observation not in line with the rest of the data. The prompt detection of such anomalies together with a suitable visualisation of these patterns may support the identification of discrepancies in inspections' findings. Further, if required, it enables corrective actions early on in the verification process of nuclear materials.

The analytical procedure is iterative in nature. When a new set of NDA measurements is available (i.e. at the end of an inspection period), the proposed 3-steps approach will make use of past and current data concerning: (i) the Assayed Mass; (ii) the uncertainty of the Assayed Mass obtained through propagation of uncertainties due to counting statistics and calibration parameters (also known as bottom-up approach); and (iii) the Declared Mass.

It is important to remark that, even if the values of NDA measures have a temporal order, they are not analysed as time series because the time interval between measures is not constant. The quantitative methods proposed hereafter are based on time series analysis procedures with adaptations to the particular context. Moreover, the possible presence of outliers in the data requires the adoption of "robust" statistical methods, i.e. methods that allow obtaining estimates that are not affected by anomalous observations (the outliers) or other deviations from model assumptions. For a detailed and comprehensive description of robust statistics, see (inter alia) Maronna et al. (2019).

In the following sections, we describe the rationale and the expected outcome of each step. The three steps represent an exploratory study of possible analytical tools applied to NDA data designed to provide a solid basis for the development of a structured and sound statistical framework for the analysis of inspections' outcomes.

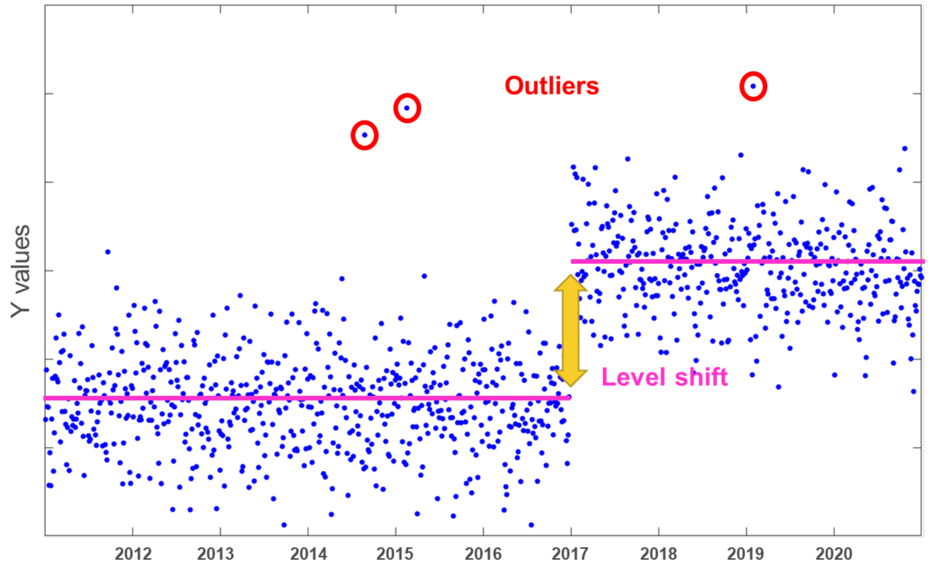

Figure 1. Example of simulated data showing the anomalous patterns that the presented approach aims to identify.

3.1. Step 1: Analysis of the ratio between the assayed mass and its uncertainty

The first variable of interest is the ratio between the Assayed Mass and its uncertainty obtained through a bottom-up approach. The assumption is that the proportion between the uncertainty and the measure is approximately constant over time. This means that the historical evolution of the ratios between each measure and its corresponding uncertainty should lie approximately on a line parallel to the *x*-axis, representing the average level of the ratios. In this context, an outlier is an Assayed Mass whose uncertainty is not coherent with the rest of measurements while a level shift is a structural break in the relation between the measured mass and its uncertainty.

In this step of the analysis, we assume that:

$$\frac{Uncertainty}{Assayed\ Mass} \sim constant \rightarrow \frac{Uncertainty}{Assayed\ Mass} = \alpha \pm \varepsilon$$

Where α represents the average level of the ratios, and ε is the error term. Therefore, when a new set of measurements is available, the first step of the statistical analysis estimates the parameters in the previous expression in a robust way to identify potential outliers and level shifts. The method adopted for the estimation is based on the robust monitoring of time series described in Rousseeuw et al. (2019). Even if conceived for time series analysis, this method can be easily adapted to the context of this paper. Involving only the Assayed Mass and its associated uncertainty obtained through a bottom-up approach, the outcome of this step provides an additional validation criterion for the measurement.

Figure 2 shows an example of the outcome of applying this step on use case 1 (Plutonium oxide cans measured in unattended mode). The analysis identifies several potential outliers, highlighted with red crosses, and a significant level shift in the average pattern of the measures. A closer look at the instrument performance revealed that in the month of May 2016 the operator modified a parameter in the measurement sequence without notification and the parameters of the algorithm triggering the data acquisition were not able to handle correctly the changes.

The impact of this modification led to a slightly higher estimation of the uncertainty of the assayed mass from 1.10% to 1.30% on average (one relative standard deviation, expressed in percentage).
Therefore, the statistical analysis proves to be able to detect changes in the measurement conditions, providing a timely indication for inspectors and technicians to plan for any remedial actions (i.e. an instrument's maintenance operation).

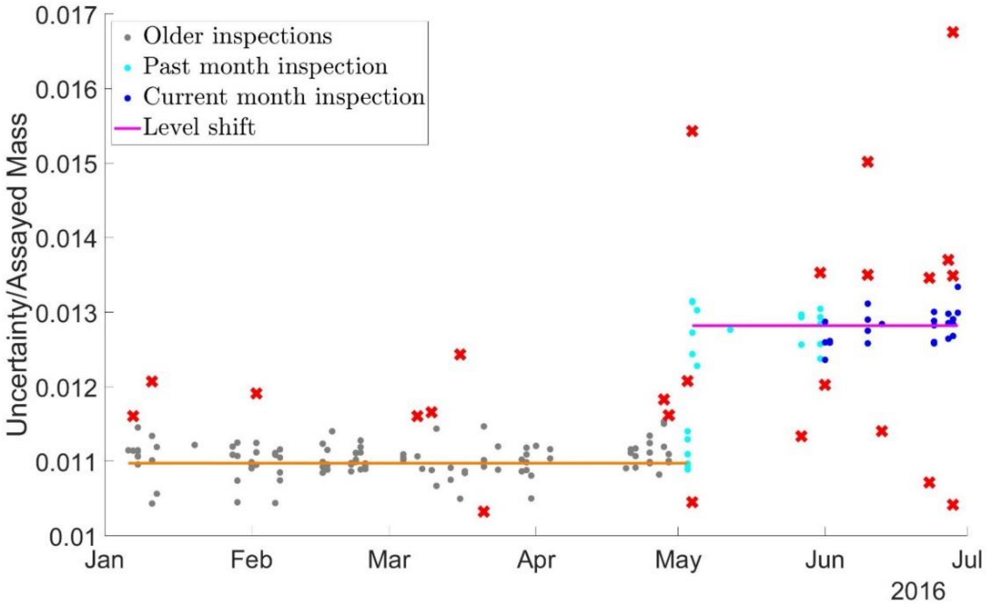

Figure 2. Outcome of the first step of the analysis on use case 1 (Pu oxide cans measured in unattended mode). The level shift indicates a structural break in the relation between the measured mass and its uncertainty, while the red crosses represent potential outliers (i.e. assayed masses whose uncertainties are not coherent with the rest of measurements).

The assumption of a constant relation between mass and uncertainty for a given instrument does not always reflect the reality of the measurement conditions. This especially holds when an instrument is used to measure a wide range of masses. In this case, the uncertainty is affected by the mass for both the counting statistics (higher masses generally produce higher count rates), and for the calibration parameters. This results in a higher dispersion of the points around the average value α and a higher error term ε which, in turn, leads to a less accurate definition of outliers and more potential missed 'alarms' (false negatives). Nevertheless, the advantage of the proposed approach is that it can be applied in the exact same way for any type of measurement or instrument, making it a robust and consistent approach for any safeguards measures. On the contrary, the adaptation of the approach for any individual function

$$\frac{Uncertainty}{Assayed\ Mass} \sim f(Assayed\ Mass) \pm \varepsilon$$

would not only complicate the approach, but also make it extremely sensitive to changes in the measurement conditions and thus subject to frequent adjustments.

3.2. Step 2: Identification of systematic bias and outliers in the sequence of relative Operator-Inspector differences

The second variable of interest is the relative Operator-Inspector Difference, defined as:

$$ROID = \frac{Assayed\ Mass - Declared\ Mass}{Declared\ Mass}$$

This variable is also expected to be approximately constant over time, that is:

$$ROID = \beta \pm \epsilon$$

where β represents the average level of the ROIDs, and ϵ is the error term. This expression being very similar to the one introduced in the previous section, it will be analysed through the same statistical approach. In addition to the identification of potential outliers and level shifts, in this case we are also interested in a careful assessment of the coefficient β. This represents the average relative distance between the operator and inspector measurements. An estimate of β significantly different from zero suggests there is systematic deviation between the two measures.

Figure 3 presents data from use case 2 (impure Pu items measured in attended mode) where the estimate of β appears to be not statistically different from zero. Figure 4 shows an example obtained with use case 1 (Pu oxide cans measured in unattended mode) where a systematic bias between the operator and the inspector measurements is visible. Finally, Figure 5 presents an additional example from use case 1 where the bias, in addition to being systematic, also increases significantly over time. In all three cases, several potential outliers are identified.

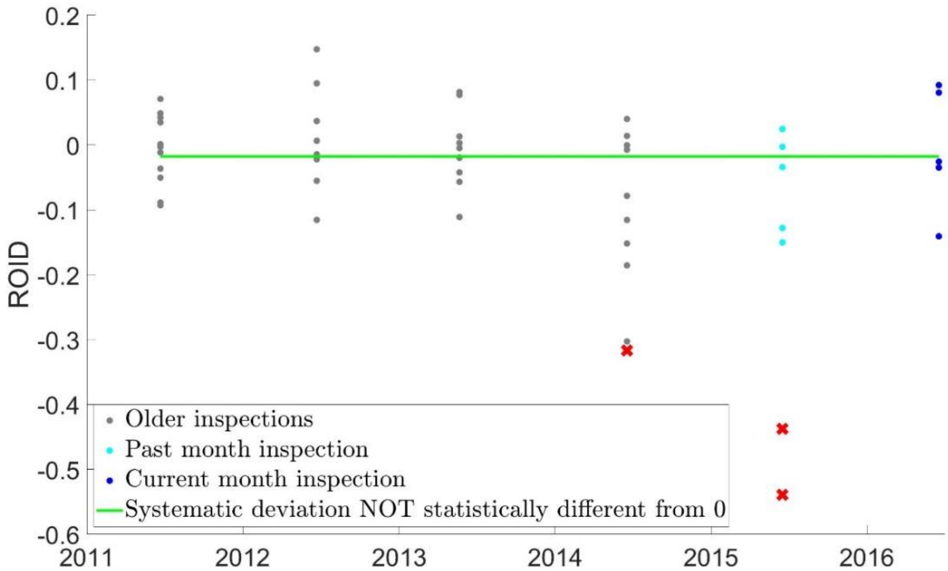

Figure 3. Outcome of the second step of the analysis on use case 2 (impure Pu items measured in attended mode). The estimate of β represents the average value of the Relative Operator-Inspector Difference (ROID) expressed in percentage that appears to be not statistically different from zero. The red crosses indicate potential outliers whose ROIDs are not coherent with the rest of the measurements.

The presence of a systematic bias between the verification measurements and the operator declarations is undesirable, but sometimes it is quite unavoidable. For example, in an unattended measurement station it can be due to an erroneous background measured when the nuclear material bearing item is too close to the instrument. The bias could be reduced by a change in the measurement sequence, but that may be very difficult to implement for the operator. In such cases, once the bias is recognised, understood and accepted, it is important to take it into account in the following evaluation of the measurement.

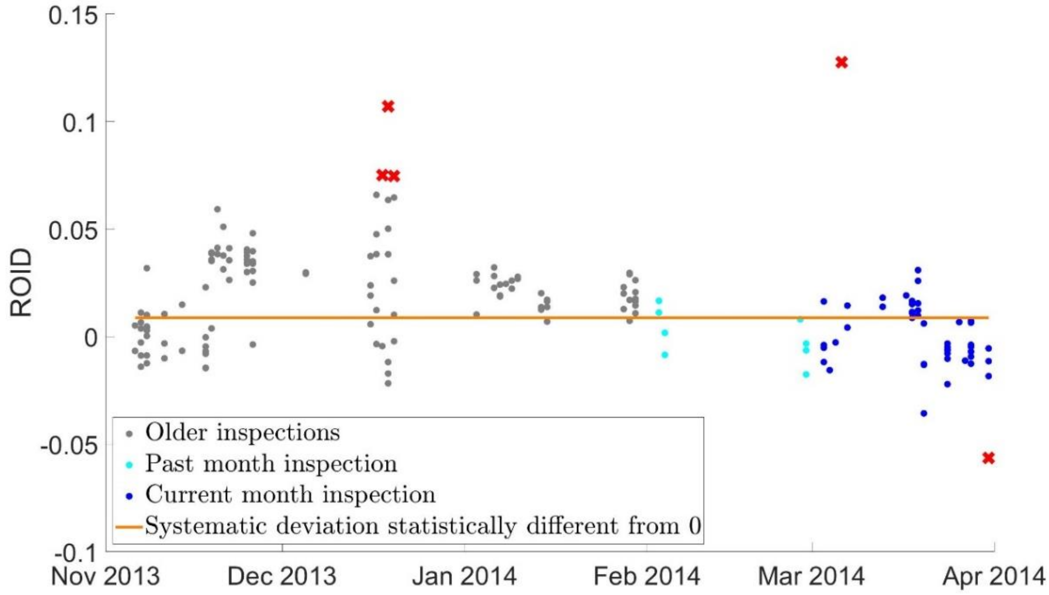

Figure 4. Outcome of the second step of the analysis on a subset of use case 1 (Pu oxide cans measured in unattended mode). The y-axis represents the Relative Operator-Inspector Difference (ROID) expressed in percentage. Here a systematic bias between the operator and the inspector measurements is visible. The red crosses indicate potential outliers.

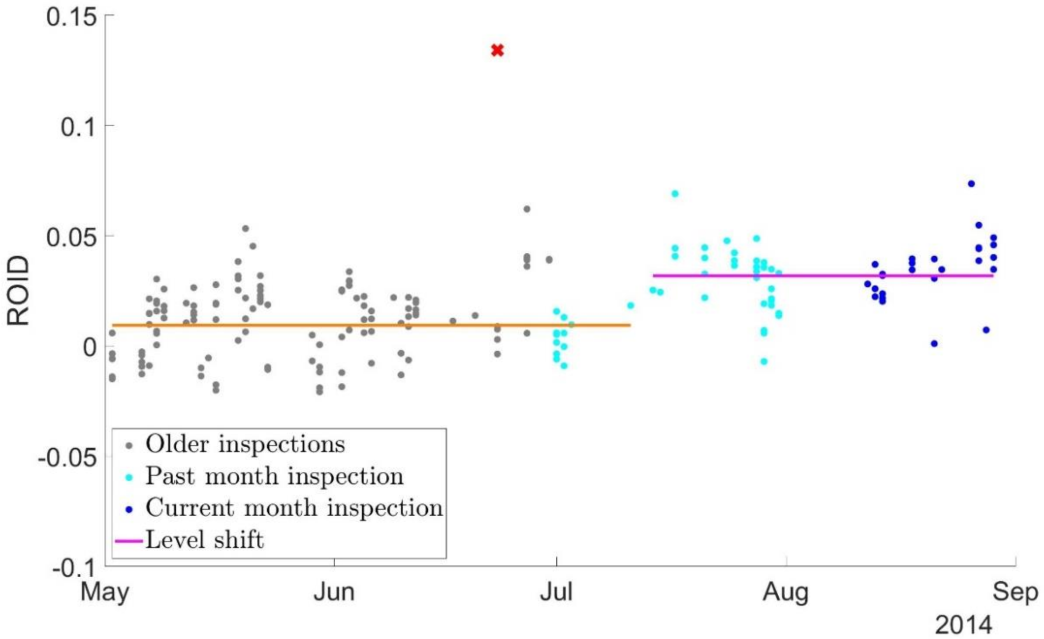

Figure 5. Outcome of the second step of the analysis on a subset of use case 1 (Pu oxide cans measured in unattended mode). The y-axis represents the Relative Operator-Inspector Difference (ROID) expressed in percentage. Here the bias is different from zero and also presents a level shift. The red crosses indicate potential outliers.

3.3. Step 3: Direct comparison of the Assayed and the Declared Mass of each item

Once the coherence of the uncertainty estimation is analysed (first step) and any systematic biases are identified (second step), the Assayed and Declared Masses are compared in the third step to verify if they are statistically different or not.
Assuming that in the previous step no systematic bias was detected (i.e. $\beta = 0$), if the Assayed Mass AM_i and the Declared Mass DM_i of an item i are not statistically different we have:

$$E(OID_i) = E(DM_i - AM_i) = 0 \qquad Var(OID_i) = Var(DM_i - AM_i) = \sigma_{i,DM}^2 + \sigma_{i,AM}^2$$

where OID is for Operator-Inspector Difference. The variance of AM_i is given by the bottom-up estimation of the uncertainty, whereas the variance of DM_i is unknown. The International Target Values (ITV) provide cut-off values η_{ITV} such that $\sigma_{i,DM} \leq \eta_{ITV} \times DM_i$. Therefore, the cut-off multiplied by DM_i represents the maximum value of the uncertainty of the Declared Mass. Along the same line, we can define:

$$\sigma_{i,DM}^2 = \eta_i^2 \times DM_i^2$$

and obtain:

$$Var(OID_i) = \eta_i^2 DM_i^2 + \sigma_{i,AM}^2$$

If we base the identification of outliers on the classical $k\sigma$ rule, where the value of k depends on the desired significance level for identifying an outlier, then an OID is an outlier if the distance from its expected value is larger (in absolute value) than k times its standard deviation. Therefore an OID is **NOT** an outlier if:

$$|OID_i - E(OID_i)| \leq k \times \sigma_{OID_i}$$

that, considering the previous expression for the variance of OID_i, becomes:

$$|OID_i - E(OID_i)| \leq k\sqrt{\eta_i^2 DM_i^2 + \sigma_{i,AM}^2}.$$

This condition is satisfied whenever

$$\eta_i \geq \frac{1}{kDM_i}\sqrt{(OID_i - k\sigma_{i,AM})(OID_i + k\sigma_{i,AM})}.$$

Therefore, the Operator-Inspector Difference of an item i is **NOT** an outlier whenever the ratio between the uncertainty and the value of Declared Mass (represented by η_i) is above a value $\hat{\eta}_i$, given by:

$$\hat{\eta}_i = \frac{1}{kDM_i}\sqrt{(OID_i - k\sigma_{i,AM})(OID_i + k\sigma_{i,AM})}.$$

The higher the value of $\hat{\eta}_i$, the higher the probability that OID_i is an outlier. If the value under the squared root is negative, there is no statistical evidence to identify OID_i as an outlier, according to the chosen outlier definition strategy. Moreover, we can directly compare $\hat{\eta}_i$ with the ITV cut-off η_{ITV}. This yields to more interpretable results, and simplifies the detection of suspiciously high values of $\hat{\eta}_i$. In particular we can distinguish 3 cases:

1. $\hat{\eta}_i = 0$ (because the value under the square root is smaller than 0): no statistical evidence of anomaly;
2. $0 < \hat{\eta}_i \leq \eta_{ITV}$: in this case we need to identify which values of $\hat{\eta}_i$ are suspiciously high. However, the assessment of the magnitude of $\hat{\eta}_i$ should be easily defined, given its direct connection with η_{ITV}.

3. $\hat{\eta}_i > \eta_{ITV}$: the Assayed and the Declared Mass are statistically different, we are in presence of an outlier.

Finally, we started this section by assuming that in the analysis of step 3.2 no systematic bias was detected (i.e. $\beta = 0$). If this is not the case, then we have that:

$$E(OID_i) = E(DM_i - AM_i) = \beta DM_i \qquad Var(OID_i) = Var(DM_i - AM_i) = \sigma_{i,DM}^2 + \sigma_{i,AM}^2.$$

Therefore the OID is **NOT** an outlier if:

$$|OID_i - \beta DM_i| \le k\sqrt{\eta_i^2 DM_i^2 + \sigma_{i,AM}^2}$$

that leads to the following expression for the threshold:

$$\hat{\eta}_i = \frac{1}{kDM_i}\sqrt{(OID_i - \beta DM_i - k\sigma_{i,AM})(OID_i - \beta DM_i + k\sigma_{i,AM})}.$$

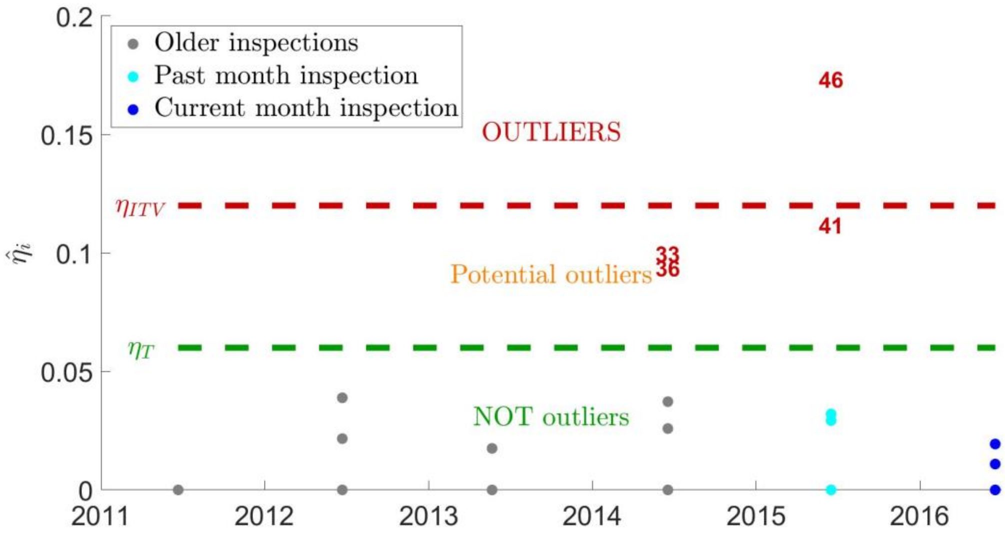

Figure 6. Outcome of the third step of the analysis on use case 2 (impure Pu items measured in attended mode). The y-axis represent the variable $\hat{\eta}_i$, a new evaluation parameter based on the ratio between the uncertainty and the value of Declared Mass. The value η_{ITV} define the threshold between suspiciously high values of $\hat{\eta}_i$ and obvious outliers, whereas η_T discriminates the "safe" values of $\hat{\eta}_i$ from the ones that may raise an alarm. At the end of the analysis, the measurements identified as outliers are reported by their measurement ID number, allowing inspectors for a quick graphical overview of the anomalous measurements.

Figure 6 presents an example from use case 2 (impure Pu items measured in attended mode) of the application of this new threshold. The values of η_{ITV} and η_T, that divide the plot into three sub-regions, are purely descriptive, in order to demonstrate how the outcome of this approach may be interpreted. The former define the border between suspiciously high values of $\hat{\eta}_i$ and obvious outliers, whereas the latter discriminates the "safe" values of $\hat{\eta}_i$ from the ones that may raise an alarm. This particular case is an interesting example as the items were shipped between two facilities and measured at both the shipper and the receiver with different instruments. A cross check between the results at both ends led to a confirmation that measurements number 41, 46, 33 and 36 were inconsistent with the declarations and that the corresponding OIDs were actually outliers.

4. Conclusion and future work

The paper presents a proposal for a statistical analysis aimed at identifying trends in historical NDA measurements data. The proposal was recently presented to Euratom Safeguards inspectors and NDA technicians receiving positive comments. Specifically, inspectors recognised that the statistical analysis would complement the evaluation of data by calculation of acceptance criteria based on historical results, while NDA technicians found it a useful tool for monitoring equipment performance over time, planning maintenance intervention accordingly and timely detecting equipment failures.
However, further analysis is required with larger sets of representative inspection's measurements to finalise the statistical approach. There is a variety of NDA systems deployed in field, each with its own specificities, some are even uniquely custom made for a single application, and the statistical analysis needs to be able at the same time to adapt to measurement specificities while applying a consistent approach throughout material balance areas, installations and even throughout countries.
In order to accomplish this task to the best possible extent, the practical implementation of the tool foresees to couple the statistical analysis with additional validation criteria based on a set of parameters defined a-priori that take into account the very detailed feature of each NDA instrumentation. This would add another layer of confidence in instrument performance that would timely alert the inspectors in their evaluation of the verification measurement. Moreover, the validation criteria would prevent the loading of erroneous results in the statistical analysis (for example, results containing clerical errors in the definition of the measurement parameters, some of which are easily overlooked during intense inspection activities). This, coupled with the application of a robust statistical method described earlier, is a further assurance that the evaluation criteria derived by the approach proposed will likely not be affected by anomalous observation.

5. References

Official Journal of the European Union C 327, 26.10.2012, Consolidated version of the Treaty establishing the European Atomic Energy Community

Maronna, R. A., Martin, R. D., Yohai, V. J., and Salibian-Barrera, M. (2019). Robust statistics: theory and methods (with R). John Wiley & Sons.

Rousseeuw, P., Perrotta, D., Riani, M., and Hubert, M. (2019). Robust monitoring of time series with application to fraud detection. Econometrics and statistics, 9:108-121.

K. Zhao, M. Penkin, C. Norman, S. Balsley, K. Mayer, P. Peerani, C. Pietri, S. Tapodi, Y. Tsutaki, M. Boella, G. Rehna Jr., E. Kuhn, International Target Values 2010 for Measurement Uncertainties in Safeguarding Nuclear Materials, ESARDA Bulletin, No. 48, December 2012.